

ПРИКЛАДНОЕ ЯЗЫКОЗНАНИЕ

ОБЩИЕ ПРИНЦИПЫ МАШИННОГО ПЕРЕВОДА С КИТАЙСКОГО ЯЗЫКА

Основная задача машинного перевода с современного китайского языка состоит в переработке той грамматической информации, которую дает в наше распоряжение иероглифический текст.

Для человека, не знающего китайского языка, китайский текст представляет собой цепь знаков, находящихся на равном расстоянии друг от друга. Через некоторые промежутки эта цепь прерывается более широкими интервалами, куда помещены те или иные знаки препинания. Следует отметить, что в современном китайском языке интервалы между иероглифами, служащими для обозначения целых слов, и иероглифами, служащими для обозначения только части слова (морфемы), одинаковы. Вот почему первой практической задачей машинного перевода с китайского языка будет разделение текста на отдельные слова. Для этого необходимо применение автоматического устройства, способного делить цепь знаков на отдельные слова (принципы работы подобного устройства будут изложены в отдельной статье). Здесь мы рассмотрим только грамматические принципы перевода с современного китайского языка на русский, причем предлагаемые принципы имеют значение в первую очередь для переводов естественнонаучной и технической литературы и не учитывают целого ряда грамматических явлений, весьма широко распространенных в языке художественных произведений.

Современный китайский язык представляет собой наиболее совершенный образец языка аналитического строя, где все возможные грамматические отношения между словами в предложении могут быть выражены только двумя способами: служебными словами и порядком знаменательных слов. Именно это обстоятельство должно быть принято во внимание при составлении правил перевода с помощью машины. Таким образом, знаки препинания, служебные слова и порядок знаменательных слов послужат источником грамматической информации, которую должна переработать машина.

Прежде всего необходимо выяснить, в каких отношениях друг к другу могут находиться знаменательные слова китайского языка, когда между ними отсутствуют служебные элементы. Известно, что по законам синтаксиса современного китайского языка из двух примыкающих друг к другу существительных первое является определением ко второму. Если за существительным следует глагол, то такое сочетание образует фразу, где существительное выступает как подлежащее, а глагол — как сказуемое. Если существительное следует за глаголом, то в таком сочетании существительное является дополнением к глаголу. Последовательность — существительное, глагол, второе существительное — образует предложение, где первое существительное представляет собою подлежащее, глагол — сказуемое, а второе существительное — дополнение. Собственно, в современном китайском языке возможны только эти четыре типа грамматических отношений между знаменательными словами.

В данной статье отрезки предложения, в которых отношения между словами определяются только их порядком, мы будем называть синтагмами. Естественно, что четырем типам грамматических отношений соответствуют четыре типа синтагм.

1. Именные синтагмы

Именную синтагму может образовать отдельное существительное или прилагательное¹.

S (1)
A (1')

¹ Здесь и далее приняты следующие обозначения: S — существительное, A — прилагательное, V — глагол, Vm — модальный глагол, Vc — связка, Ad — наречие, P — местоимение.

Именная синтагма может быть образована сочетанием двух существительных или прилагательного с существительным:

S'S (2)
AS (2')

Синтагмы (2) и (2') могут выступать как определение к S, т. е. к (1):

S'S'S (3)
AS'S (3')

И наоборот, синтагма (1) может выступать в роли определения к S'S и AS, т. е. к (2) и (2').

S'S'S (4)
S'AS (4')

Легко заметить, что синтагмы (3) и (4), формально не отличающиеся друг от друга, в действительности отличаются группировкой членов. Внешне это создает неразрешимый случай для машины, и может показаться, что для таких синтагм придется давать два варианта перевода. Однако при переводе на русский язык эта разница исчезнет, поскольку как в первом, так и во втором случае первые два члена будут переведены существительными в родительном падеже.

Синтагмы (2) и (2') могут также сочетаться друг с другом и сами с собой:

S''S'S'S (5)
A'S'AS (5')
S''S'AS (5'')
AS''S'S (5''')

Теоретически процесс сложения можно продолжить сколько угодно, однако практически количество слов в рассматриваемых синтагмах редко превышает четыре. Обычно цепь определений бывает прервана определительно-именным суффиксом -ды, в результате чего она распадается на две, а иногда и на три синтагмы меньшего размера.

Представителем любой из синтагм (2) — (5) может служить их центральное слово — существительное S, стоящее на правом краю синтагмы. В последующих схемах оно везде будет употребляться только для обозначения отдельного существительного, однако при этом следует иметь в виду, что в действительности там может иметь место любая синтагма (2) — (5).

2. Глагольные синтагмы

V (6)
(Ad) VS (7)

3—4. Предикативные синтагмы

S (Ad) V (8)
S (Ad) A (9)
S(Ad) VS' (10)

Возможны также и варианты синтагм. Так, при наличии модальных глаголов синтагмы (6) — (10) примут следующий вид:

(Ad) Vm (Ad) V (6')
(Ad) Vm (Ad) VS (7')
S (Ad) Vm (Ad) V (8')
S (Ad) Vm (Ad) A (9')
S (Ad) Vm (Ad) VS' (10')

При помощи тех или иных служебных элементов синтагмы соединяются в целое предложение. При этом в зависимости от характера связующего служебного элемента возможны два типа связи между синтагмами: согласование и примыкание. В первом случае грамматическая форма центрального слова синтагмы согласована с грамматической формой того или иного слова, находящегося в другой синтагме; во втором случае грамматическая форма центрального слова определяется только отношениями между словами внутри данной синтагмы.

Каждая из приведенных выше синтагм имеет определенный типовой перевод на русский язык.

Так, синтагма (1) переводится следующим образом: существительное переводится

в зависимости от характера согласования [за исключением S, входящих в правую часть синтагм (7), (7'), (10), (10'), которые переводятся в зависимости от характера глагольного управления]. Синтагма (1') переводится прилагательным, которое согласуется с существительным соседней синтагмы.

Синтагма (2) переводится: S — в зависимости от характера согласования, S' — род. падеж соответствующего числа. Перевод члена S' ставится на второе место.

Синтагма (2') переводится: A — прилагательным, согласованным в роде и числе с S. Порядок слов в переводе не изменяется.

Синтагма (3) переводится: S — в зависимости от характера согласования, S'' и S' — род. падеж. Перевод члена S''S' ставится после перевода члена S.

Синтагма (3') переводится: S — в зависимости от характера согласования, S' — род. падеж. Прилагательное A согласуется в роде, числе и падеже с S'. Перевод члена AS' ставится на второе место.

Синтагма (4) переводится так же, как и синтагма (3).

Синтагма (4') переводится: S — в зависимости от характера согласования. Прилагательное A согласуется с S. Существительное S' — род. падеж. Перевод члена S' ставится после перевода члена AS.

Синтагма (5) переводится: S — в зависимости от характера согласования. Существительные S''S''' и S' — род. падеж. При переводе члены синтагмы выстраиваются в следующем порядке: SS'S''S'''.

Синтагма (5') переводится: AS — как в синтагме (2'), A'S' как в синтагме (3'). Перевод члена A'S' ставится на второе место.

Синтагма (5'') переводится: AS — как в синтагме (2'), S'S' — как в синтагме (3). Перевод члена S'S' ставится на второе место.

Синтагма (5''') переводится: S'S — как в синтагме (1), AS'' — как в синтагме (3'). Перевод члена AS'' ставится на второе место.

В тех случаях, когда рассмотренные выше именные синтагмы представляют собой правую часть предикативных синтагм (7), (7'), (9), (9'), грамматическая форма центрального слова определяется характером глагольного управления.

Синтагма (6) переводится глаголом в соответствующей форме.

Синтагма (7) переводится: глагол V — как в синтагме (6), а S — как в синтагмах (1) — (5) в соответствии с характером управления глагола.

Синтагма (8) переводится: существительное S — как в синтагме (1) — (5), V согласуется с S.

Синтагма (9) переводится: S — как в синтагмах (1) — (5), V согласуется с S, а S' переводится в зависимости от характера глагольного управления V.

По своему назначению для машинного перевода служебные элементы современного китайского языка можно разделить на две группы: служебные элементы, при помощи которых осуществляется разделение предложения на отдельные синтагмы, и служебные элементы, помогающие определить принадлежность знаменательных слов к определенной части речи.

Точками разграничения синтагм в первую очередь служат знаки препинания. Система пунктуации современного китайского языка основана на чисто логических принципах, и поэтому запятая не может стоять между элементами синтагм, представляющих собой замкнутые целые как в логическом, так и в грамматическом отношении.

Точками разграничения синтагм безусловно послужат различные глаголы-предлоги: *гэй* «для, ради», *ю* «посредством, при помощи», *дуй* «относительно», *цун*, *ю* «из», *цзы*, *цзыцун* «с тех пор как», *дао* «к, в, до», *цзай* «в», и т. п. Глаголы-предлоги открывают синтагму, и эта синтагма продолжается вплоть до ближайшего знака препинания или (при глаголе-предлоге *цзай*) до послелога.

В современном китайском языке широко распространены так называемые рамочные конструкции, служащие для обозначения разного рода пространственных или временных отношений. Начальным элементом рамочной конструкции служит глагол-предлог *цзай* «в», «находиться в...» или служебный глагол *дан* «быть», а замыкающим — послелог или существительное со значением места или времени. Примерами таких конструкций могут служить следующие: *цзай... чжун* «в, внутри», *цзай... шан* «на», *цзай... ся* «под», *цзай... цзянь* «перед», *цзай... хоу* «за», *цзай... цзянь* «между», *цзай... вай* «вне», *цзай(дан)... ши* «когда, во время» и т. п.

В такой рамочной конструкции может быть заключена любая из перечисленных выше синтагм, но чаще всего встречаются именные синтагмы.

Различные типы синтагм, рассмотренные выше, имеют формальные границы справа и слева. Однако синтагмы с определительно-именным суффиксом *-ды* могут иметь только одну формально выделяемую границу справа — сам суффикс. Левая граница таких синтагм определяется по-разному. При этом возможны два варианта.

1. Определительная синтагма стоит в начале предложения или в начале синтагмы, формально ограниченной слева. В этих случаях левой границей определительной синтагмы служит начало предложения или начало более крупной синтагмы.

2. Определительная синтагма стоит в конце предложения, после глагола. В этом случае левой границей определительной синтагмы является не служебный, а знамена-

тельный элемент — глагол, который может быть обнаружен только после определения принадлежности слов к частям речи. Поскольку определительная синтагма всегда помещается только внутри большой синтагмы, имеющей формальные границы справа и слева, назовем ее субсинтагмой.

Когда к рассмотренным выше синтагмам присоединяется суффикс *-ды*, их перевод изменяется следующим образом.

Синтагма (1) переводится: существительное S ставится в родительном падеже, или, если имеется соответствующее относительное прилагательное, употребляется это прилагательное.

Синтагма (1'). Прилагательное A согласуется с существительным соседней синтагмы справа.

Синтагмы (2) — (5). Центральное слово таких синтагм — существительное — ставится в родительном падеже.

Синтагма (6) переводится действительным или страдательным причастием, в зависимости от характера глагола.

Синтагма (7) переводится причастным оборотом или определительным придаточным предложением.

Синтагмы (8) и (10) переводятся определительными придаточными предложениями.

После выяснения границ синтагм наиболее важной задачей является определение принадлежности к частям речи слов, входящих в синтагму. Установить принадлежность слова к той или иной части речи в современном китайском языке не всегда просто из-за частого явления конверсии. Наиболее распространенный случай конверсии: глагол — существительное; менее распространенный: существительное — прилагательное. Однако грамматическое оформление и синтаксическое окружение резко снижают возможности двусмысленного определения принадлежности слова к той или иной части речи. Служебные и знаменательные элементы, указывающие на принадлежность слов к определенным частям речи, — это слова-индикаторы. К индикаторам глагола в первую очередь относятся суффиксы видо-временного оформления¹: суффикс настоящего длительного времени *-и жеэ*, суффикс прошедшего совершенного времени *-ла*, суффикс давнопрошедшего времени *-го*. Эти грамматические элементы указывают, что слово, стоящее влево от него, является глаголом². В эту же группу входят и наречия. В современном китайском языке наречия всегда занимают место перед глаголом или глаголом-предлогом. Таким образом, наречия *ду* «все», *ю* «снова», *е* «также», *хай* «все еще», *и* «уже» и т. п. указывают, что слово, стоящее вправо от них, является глаголом.

Таким образом, в современном китайском языке имеется достаточное количество формальных показателей, при помощи которых можно установить принадлежность того или иного слова к категории глагола. Это особенно важно в связи с тем, что в языке естественно-научной и технической литературы, которая в первую очередь делается объектом машинного перевода, глагольные сказуемые в подавляющем большинстве предложений имеют один из перечисленных видов оформления.

Существительное не имеет столь удобных опознавательных знаков, свойственных глаголам. Однако некоторую помощь в этом отношении могут оказать счетные слова: *зеэ* — счетное слово для предметов вообще, *тао* — для предметов удлиненной формы, *цзя* — для механизмов, приборов и т. п. Счетное слово с предшествующим ему числительным или указательным местоимением далеко не всегда стоит непосредственно перед существительным, к которому относится. Если перед существительным имеется определение, счетное слово ставится перед определением³, иначе говоря, оно ставится перед синтагмами (1) — (5) в целом.

Счетные слова с предшествующими им числительными и указательными местоимениями могут оказать существенную помощь при определении числа существительного⁴. Существительные современного китайского языка, как известно, обозначают не единичные предметы, а всю совокупность предметов, имеющих одинаковое название, поэтому, естественно, они не имеют формальных показателей множественного числа. Числительные или указательные местоимения со счетными словами как бы выделяют из этой совокупности единичные предметы. Поэтому существительные, не имеющие при себе счетных слов, можно безошибочно переводить множественным числом (за исключением слов, обозначающих единичные понятия типа *цзунго* «родина», *Чжунго* «Китай» и некоторых других). Напротив, число существительных, имеющих счетные слова, должно переводиться в зависимости от числительного, стоящего перед счетным словом.

¹ Напомним, что интервал между основной и суффиксом на письме не отличается от интервала между двумя знаменательными словами.

² Это справедливо по крайней мере для языка естественно-научной и технической литературы.

³ По крайней мере, в языке естественно-научной и технической литературы.

⁴ В тех случаях, когда счетное слово стоит перед синтагмами (1) — (5), оно относится только к центральному слову синтагмы, стоящему на последнем месте.

Формальным признаком наречий, образованных от прилагательных, служит суффикс *-ди*, который фонетически совпадает с определительно-именным суффиксом *-ды*, но на письме отличен от него.

Здесь перечислены наиболее часто встречающиеся служебные слова современного китайского языка, которые могут служить индикаторами. При составлении подробных правил перевода количество индикаторов можно расширить за счет других менее распространенных служебных и знаменательных элементов. Однако имеются случаи, когда одно или несколько слов в синтагме не могут быть отнесены к какой-либо определенной части речи.

Наибольшие трудности возникают при переводе предикативных синтагм. Главной задачей перевода таких синтагм является отыскание глагола, представляющего собой центральный член. В современном китайском языке этот глагол в подавляющем большинстве случаев бывает оформлен. Если же окажется, что единственный глагол синтагмы не имеет никакого оформления, что довольно трудно себе представить, то это означает неразрешимый случай, при котором машина выдает несколько возможных вариантов перевода.

При переводе рамочных конструкций следует иметь в виду, что в рамочных конструкциях со значением места обычно бывают заключены именные синтагмы, а в рамочных конструкциях со значением времени — предикативные. Следовательно, все альтернативы $\frac{S}{V}$ и $\frac{V}{S}$ решаются в пользу S , а альтернативы $\frac{S}{A}$ и $\frac{A}{S}$ разрешаются после сопоставления с типовыми конструкциями именных синтагм. Допустим, принадлежность слов именной синтагмы к частям речи определяется следующим образом: $\frac{S}{V} \frac{A}{S}$. Сразу же устраняется первая альтернатива, поскольку в именных синтагмах глагол невозможен, остается $S \frac{A}{S}$. Последовательность SA невозможна и поэтому отсутствует в списке именных синтагм. Следовательно, остается SS .

Наиболее сложным представляется перевод именных субсинтагм. Сложность состоит в том, что, помимо трудностей отыскания левой границы субсинтагмы, структура ее никогда не может быть заранее определена. В связи с тем, что глагол в определительной глагольной или предикативной субсинтагме, как правило, не имеет грамматического оформления, его часто невозможно отличить от существительного. Поэтому в целом ряде определительных субсинтагм будут иметь место неразрешимые случаи, для которых придется давать несколько вариантов перевода.

Машинный перевод с китайского языка будет, по-видимому, совершаться в следующей последовательности: 1) расчленение предложения на синтагмы; 2) установление принадлежности слов к определенным частям речи; 3) выделение субсинтагм; 4) перевод синтагм; 5) объединение переведенных синтагм в предложение русского языка. Рассмотрим эти ступени в отдельности.

1. Первой ступенью машинного перевода является расчленение предложения (т. е. цепи знаков, заключенных между двумя точками или началом абзаца и точкой) на синтагмы в соответствии с изложенными выше принципами. Возьмем предложение: 科學工作在祖國的社會主義建設事業中担负的任務是繁重的,這繁重的程度還將一年比一年增加. *кэсюэ гунцзо цзай цзугу-ды цзхуйчжун цзяньши ши чжун даньфу-ды жэньгу ши фаньчжун-ды, чже-аньчжун-ды чэнду хай цзян и-нянь-би-и-нянь цзэнцзя*. Из предложения выделяется рамочная конструкция со значением места (*цзай... чжун*): 在祖國的社會主義建設事業中 *цзай цзугу-ды цзхуйчжун цзяньши ши чжун*. Остальную часть запятая делит на две синтагмы: 科學工作担负的任務是繁重的 *кэсюэ гунцзо даньфу-ды жэньгу ши фаньчжун-ды* и 這繁重的程度還將一年比一年增加. *чжэ фаньчжун-ды чэнду хай цзян и-нянь-би-и-нянь цзэнцзя*. Одновременно выделяются и служебные слова — индикаторы, указывающие на принадлежность слов к той или иной части речи. В нашем примере наречие *還 хай* указывает, что следующее за ним слово *將 цзян* является глаголом¹; идиоматический оборот *一年比一年 и-нянь-би-и-нянь*, который синтаксически проявляет себя как наречие, тоже показывает, что следующее за ним слово *增加 цзэнцзя* является глаголом, а не существительным.

2. Следующей ступенью является определение принадлежности слов к тем или иным частям речи:

Первая синтагма: $S-ды S \frac{S}{V} S$

Вторая синтагма: $S \frac{S}{V} V-ды SV \frac{S}{A}-ды$

Третья синтагма: $P \frac{S}{A}-ды S Ad Vm Ad V$

¹ *Цзян* может быть как модальным глаголом, так и глаголом-предлогом. Если после *цзян* стоит существительное, то это безусловно глагол-предлог; если же после него стоит глагол, то это модальный глагол.

В первой синтагме это определение будет двусмысленным. Однако известно, что в рамочной конструкции со значением места глагол невозможен, следовательно альтернатива $\frac{S}{V}$ решается в пользу S. В окончательном виде первая синтагма выглядит следующим образом: S-ды SSS.

Во второй синтагме определение также двусмысленно. Однако известно, что в усилительной конструкции возможно только A. Кроме того, известно, что в синтагме может быть только один глагол, следовательно альтернатива в субсинтагме $S \frac{S}{V} V$ -ды должна быть разрешена в пользу S. В окончательном виде вторая синтагма выглядит следующим образом: SSV -ды SvCa -ды.

В третьей синтагме снова имеет место альтернатива. Однако суффикс -ды показывает, что эту альтернативу следует разрешить в пользу A. В окончательном виде третья синтагма выглядит следующим образом: PS-ды S Ad Vm Adv.

3. На третьей ступени выделяются определительные субсинтагмы. Приведенный пример не совсем удачен для того, чтобы показать, почему определительные субсинтагмы следует выделять после определения принадлежности слов к тем или иным частям речи. В этом примере определительные субсинтагмы находятся в левой части синтагм, поэтому их левые границы совпадают с левыми границами синтагм. Однако если бы они находились в правой части синтагмы, их левой границей послужил бы глагол, т. е. не формальный, а знаменательный элемент, который сам по себе нуждается в определении.

После выделения субсинтагм получаем синтагмы, в которых полностью отсутствуют служебные слова, и все грамматические отношения между словами определяются только их порядком.

祖國的	цзуго-ды	синтагма (1)
科建設事業中在社會主義	цзай шэжуйчжун цзянь- шэ шие чжун	синтагма (3)
學工作担負的	кэжкэ гуицзо дан фу-ды	синтагма (8)
任務是繁重的	жэньу ши фаньчжун-ды	синтагма (10)
程度還將一年比一年增加。	чэндю хай цзян и-нянь би-и-нянь цзэцзэя	синтагма (8)

4. По изложенным выше правилам осуществляется перевод отдельных синтагм: 1) «родины»; 2) «в деле социалистического строительства»; 3) «за которые несет ответственность научная работа»; 4) «задачи разнообразны»; 5) «этого разнообразия»; 6) «степень еще будет увеличиваться из года в год». На этой же ступени осуществляется и согласование между центральными словами синтагм.

5. Соединение переведенных синтагм проводится в обратном порядке. В первую очередь переводы определительных субсинтагм присоединяются к тем существительным, к которым относятся: 1 + 2 «в деле социалистического строительства родины»; 3 + 4 «задачи, за которые несет ответственность научная работа»; 5 + 6 «степень разнообразия еще будет увеличиваться из года в год».

Перед нами перевод синтагм, которые получились в результате расчленения предложения на первой ступени перевода. Затем переводы этих синтагм располагаются в том же порядке, как и в китайском предложении. При этом возможны некоторые отклонения, регулируемые специально введенными правилами.

В результате получается перевод: «Задачи, за которые несет ответственность научная работа в деле социалистического строительства родины, весьма разнообразны, степень разнообразия еще будет увеличиваться из года в год».

М. В. Софронов

О ЛИНГВИСТИЧЕСКОЙ ВЕРОЯТНОСТИ

Знать язык — это значит хранить в памяти слова и грамматические модели этого языка, определяемые его грамматическим строем. Слова хранятся в памяти в виде звуковых образов, а у грамотных людей — и в виде зрительных. Последние либо соотносимы со звуковыми образами, если мы имеем дело с языками, пользующимися буквенным (или, как его часто называют, звуковым) письмом, либо не соотносимы с ними, если мы имеем дело с языками, пользующимися иероглифами. Звуковой облик отдельных слов состоит из некоторого числа повторяющихся в разнообразных сочетаниях элементов, именуемых звуками речи или фонемами, составляющими в каждом языке свою особую систему.